

CLAIMS

What is claimed is:

1 1. A method, for use in query optimization in a relational database management system,
2 said method comprising the steps of:

3 (a) generating statistical information regarding data which represents the results of
4 an operation involving one or more columns of a database;

5 (b) deriving a statistical soft constraint from said statistical information that reflects a
6 statistical property of said data; and

7 (c) using said statistical soft constraint to estimate a cardinality value for the result of
8 applying one or more query predicates in a query plan.

1 2. The method of claim 1 further comprising the step, prior to step (a), of creating a
2 materialized column containing said data, wherein said data comprises the results of said
3 operation involving one or more columns of a database.

1 3. The method of claim 2 wherein said materialized column is stored in the database.

1 4. The method of claim 2 wherein said statistical soft constraint comprises a constraint
2 predicate and an associated probability value, said associated probability value reflecting the
3 percentage of rows of said one or more columns for which said constraint predicate is true.

1 5. The method of claim 2 wherein said step of generating statistical information
2 comprises gathering said statistical information regarding said data utilizing a statistics
3 gathering process provided by the relational database management system.

1 6. The method of claim 1 wherein the step of generating statistical information
2 comprises analyzing the data using an SQL statement.

1 7. The method of claim 6 wherein said SQL statement groups a selection to obtain
2 frequencies.

1 8. The method of claim 1 further comprising the step, prior to step (b), of analyzing
2 said statistical information and determining a useful subset of said statistical information
3 from which to derive said statistical soft constraint.

1 9. The method of claim 1 wherein said statistical soft constraint comprises a constraint
2 predicate and an associated probability value, said associated probability value reflecting the
3 percentage of rows of said one or more columns for which said constraint predicate is true.

1 10. The method of claim 9 wherein said query predicate comprises an expression
2 involving two different columns.

1 11. The method of claim 10 wherein the step (c) of using said statistical soft constraint
2 comprises the steps of:

3 (c1) normalizing said query predicate, if necessary, such that the right-hand side
4 of said query predicate expression comprises a constant;

5 (c2) determining whether said query predicate matches said constraint predicate;

6 (c3) setting a selectivity for said query predicate equal to said associated
7 probability value if said query predicate matches said constraint predicate; and

8 (c4) setting a selectivity boundary for said query predicate based upon said
9 associated probability value if said query predicate does not match said constraint predicate.

1 12. The method of claim 9 wherein said query predicate comprises an operation upon a
2 column.

1 13. The method of claim 12 wherein the step (c) of using said statistical soft constraint
2 comprises the steps of:

3 (c1) normalizing said query predicate, if necessary, such that the right-hand side
4 of said query predicate expression comprises a constant;

5 (c2) determining whether said query predicate matches said constraint predicate;

6 (c3) setting a selectivity for said query predicate equal to said associated
7 probability value if said query predicate matches said constraint predicate; and

8 (c4) setting a selectivity boundary for said query predicate based upon said
9 associated probability value if said query predicate does not match said constraint predicate.

1 14. The method of claim 9 wherein said query predicate comprises two predicates, the
2 first predicate involving a first column and the second predicate involving a second column,
3 said first column being a different column from said second column, wherein said constraint
4 predicate comprises an expression including said first column and said second column.

1 15. The method of claim 13 wherein the step (c1) of using said statistical soft constraint
2 comprises the steps of:

3 (c1) normalizing said constraint predicate, if necessary, to produce a normalized
4 constraint predicate wherein the left-hand side of said normalized constraint predicate
5 comprises said first column;

6 (c2) substituting occurrences of said first column in said first predicate with the
7 right-hand side of said normalized constraint predicate, such that said first predicate only
8 refers to said second column;

9 (c3) transposing said first predicate, if necessary, to produce a transposed first
10 predicate wherein the left-hand side of said transposed first predicate comprises said second
11 column; and

12 (c4) setting a selectivity or selectivity bound based upon said transposed first
13 predicate, said second predicate and statistical information regarding said second column.

1 16. A database management system comprising:

2 means for generating statistical information regarding data which represents the

3 results of an operation involving one or more columns of a database;

4 means for generating a statistical soft constraint using said statistical information;

5 and

6 means for utilizing said statistical soft constraint to estimate a cardinality value for

7 the result of applying one or more query predicates in a query plan.

1 17. The database management system of claim 16 wherein said statistical soft constraint

2 comprises a constraint predicate and an associated probability value reflecting the

3 percentage of rows of said one or more columns for which said constraint predicate is true.

1 18. The database management system of claim 17 wherein said means for utilizing

2 comprises:

3 means for identifying a type of said query predicate;

4 means for normalizing said query predicate and said constraint predicate;

5 means for comparing said query predicate with said constraint predicate;

6 means for setting a selectivity equal to said probability value when said query
7 predicate matches said constraint predicate; and

8 means for setting a selectivity bound based upon said probability value when said
9 query predicate does not match said constraint predicate.

1 19. The database management system of claim 16 wherein said query predicate
2 comprises a first and second predicate.

1 20. The database management system of claim 19 wherein said means for utilizing
2 further comprises:

3 means for generating a twin predicate from said first predicate; and
4 means for setting a selectivity or selectivity bound based upon said twin predicate,
5 said second predicate and said probability value.

1 21. A computer program product comprising:

- 2 (a) a computer readable medium;
- 3 (b) code means contained in said medium for instructing a computer to perform
4 the steps of:
- 5 (i) generating statistical information regarding data which represents the
6 results of an operation involving one or more columns of a database;
- 7 (ii) deriving a statistical soft constraint from said statistical information
8 that reflects a statistical property of said data; and
- 9 (iii) using said statistical soft constraint to estimate a cardinality value for
10 the result of a query predicate in a query plan.

1 22. The computer program product of claim 21 wherein said computer readable medium

is chosen from the group consisting of a modulated electrical signal, a modulated optical signal, a magnetic storage medium and an optical storage medium.

23. A computer readable medium containing program instructions for use in query optimization in a relational database management system, said program instructions for:

(a) generating statistical information regarding data which represents the results of an operation involving one or more columns of a database;

(b) deriving a statistical soft constraint from said statistical information that reflects a statistical property of said data; and

(c) using said statistical soft constraint to estimate a cardinality value for the result of applying one or more query predicates in a query plan.